Hybrid probabilistic wind power forecasting using temporally local Gaussian process

Yan, J., Li, K., Bai, E-W., Deng, J., & Foley, A. (2016). Hybrid probabilistic wind power forecasting using temporally local Gaussian process. *IEEE Transactions on Sustainable Energy*, *7*(1), 87 - 95. https://doi.org/10.1109/TSTE.2015.2472963

Published in:
IEEE Transactions on Sustainable Energy

Document Version:
Peer reviewed version

Hybrid Probabilistic Wind Power Forecasting Using Temporally Local Gaussian Process

Juan Yan, *Member, IEEE*, Kang Li, *Senior Member, IEEE*, Er-Wei Bai, *Fellow, IEEE*, Jing Deng, and Aoife M. Foley, *Member, IEEE*

***Abstract*—The demand for sustainable development has resulted in a rapid growth in wind power worldwide. Although various approaches have been proposed to improve the accuracy and to overcome the uncertainties associated with traditional methods, the stochastic and variable nature of wind still remains the most challenging issue in accurately forecasting wind power. This paper presents a hybrid deterministic–probabilistic method where a temporally local "moving window" technique is used in Gaussian process (GP) to examine estimated forecasting errors. This temporally local GP employs less measurement data with faster and better predictions of wind power from two wind farms, one in the USA and the other in Ireland. Statistical analysis on the results shows that the method can substantially reduce the forecasting error while it is more likely to generate Gaussian-distributed residuals, particularly for short-term forecast horizons due to its capability to handle the time-varying characteristics of wind power.**

***Index Terms*—Error analysis, forecasting, Gaussian process (GP), wind power.**

I. Introduction

RENEWABLE energy, particularly wind power is nonsynchronous due to the variable and stochastic nature of wind. Accurate forecasting is problematic resulting in power system integration issues. Wind and wind power forecasting horizons can be categorized into four main time scales, i.e., very short-term, short-term, long-term, and seasonal, for different applications with different forecasting techniques [1], [2]. Broadly, there are two approaches to wind and wind power forecasting: numerical weather prediction (NWP) methods and time-series analysis. In NWP, the starting point of a fuzzy power flow tool [3] or kernel method analysis is wind speed, then it proceeds to wind power, whereas in time-series analysis, the starting point can be both wind speed and wind power. Numerous techniques are being applied for wind and wind power forecasting including density estimation [4], quantile regression [5], [6], neural networks [7], prediction intervals [8], and extreme learning machine [9].

The uncertainty associated with the forecasting has great influence on the power systems. Many studies have been carried out on the impacts and solutions. In [10], optimal short-term energy balancing to better deal with wind power uncertainty is discussed, considering reserve, market structure, and power system infrastructure changes. It has also been demonstrated that the largest single impact of wind on system operation is from the inclusion of variance and that variance, kurtosis, and skewness together produced the error information with the lowest system cost [11]. The impact of the variability of wind power on generation planning is examined in [12] and it is shown that outcomes are highly system specific (e.g., balancing, reserve, and dispatch).

One way to undertake probabilistic forecasting is to analyze the prediction errors statistically. In [13], forecasting was made using a persistence model, and the error information was collected and analyzed with a probabilistic density function (pdf). The key finding is that the pdf of error could be modeled using a Beta function, to size energy storage and to produce a smoother wind power output. Similarly, in [14], a persistence model is applied to forecast the wind power for a single wind farm, and the forecast error distribution is analyzed with a mixed probabilistic model of Laplace distribution and Dirac delta function. The results are applied to make proper penalty to the short-term markets. In [15], it is shown that wind power forecasting residuals for NWP methods do not follow Gaussian distribution after the transformation from wind to wind power. The nonlinearity in the power curve causes non-Gaussian wind power prediction errors. So, although it is often assumed that wind power forecast error has a near-Gaussian distribution, such assumption is not appropriate with some of the models mentioned above such as persistence, NWP, and also other regression models such as artificial neural networks (ANN) and autoregressive moving average (ARMA).

Gaussian process (GP) is a new regression method for wind power prediction. In [16], a sparse online warped GP has been applied to forecast wind power via wind speed taking into account the wind uncertainties associated. The key finding is that the new method adapts to the time-varying characteristic of wind generation. In [17], the authors proposed a variant GP model based on the "moving window" and TLBO (teaching–learning-based optimization) techniques, and applied it to the forecasting of the wind power generation of whole island of Ireland. In this paper, a generic temporally local GP (TLGP) model is developed, and the impacts of two key parameters in

Manuscript received February 09, 2015; revised June 17, 2015; accepted August 19, 2015. This work was supported in part by UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/L001063/1 and Grant EP/G042594/1, and in part by the National Natural Science Foundation of China under Grant 51361130153 and Grant 61273040. Paper no. TSTE-00093-2015. *(Corresponding author: Kang Li.)*

J. Yan, K. Li, and J. Deng are with the School of Electronics, Electrical Engineering, and Computer Science, Queen's University, Belfast BT7 1NN, U.K. (e-mail: jyan03@qub.ac.uk; k.li@qub.ac.uk; j.deng@qub.ac.uk).

E.-W. Bai is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa, IA 52242 USA (e-mail: er-wei-bai@uiowa.edu).

A. M. Foley is with the School of Mechanical and Aerospace Engineering, Queen's University, Belfast BT7 1NN, U.K. (e-mail: a.foley@qub.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2015.2472963

the model, e.g., window width and state vector length, on the forecast performance are analyzed. Then, TLGP is applied to two real wind farm systems in Ireland and USA. Furthermore, two additional forecasting metrics are employed to produce a more comprehensive evaluation of the forecasting performance of TLGP in comparison with four benchmark models, namely the standard GP, persistence model, ARMA, and radial basis function (RBF) model and the probabilistic distributions of forecasting residuals for all the five methods are statistically analyzed and compared to confirm the efficacy and superiority of the proposed TLGP.

This paper is organized as follows. Section I introduces significance and state-of-the-art of probabilistic wind power forecasting. Section II describes the algorithm of standard GP and persistence model. Then, the new method TLGP is developed in Section III. Section IV illustrates the case study including the training and forecasting results of wind power in two wind farms. Then, Section V presents the error comparison of the proposed method with four benchmark models under three criteria: RMSE, MSE, and MASE. Section VI analyzes the probabilistic distribution of forecasting residuals and validates the profile by analyzing the skewness and kurtosis. Section VII concludes the paper.

II. Benchmark Models

A. Standard GP

According to the definition of GP, the new output $y_0 = f(\boldsymbol{x}_0)$ should follow one joint Gaussian distribution with available data $\boldsymbol{Y} = (y_1, y_2, \dots y_N)^{\mathrm{T}}$

$$P(y_t|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}^*, \boldsymbol{x}_0) = N\left(\boldsymbol{B}\boldsymbol{C}_{\boldsymbol{Y}}^{-1}\boldsymbol{Y}, A - \boldsymbol{B}\boldsymbol{C}_{\boldsymbol{Y}}^{-1}\boldsymbol{B}^T\right) \quad (1)$$

where $\boldsymbol{\theta}^*$ is the optimal hyperparameters, $\boldsymbol{X}$ is the corresponding set of input data, and $\boldsymbol{x}_0$ is the corresponding input vector for y_0. $\boldsymbol{C}_{\boldsymbol{Y}}$, $\boldsymbol{B}$, and A are estimated according to (2)–(4), where Φ is the covariance function and the popular square exponential covariance function is given in (5)

$$\boldsymbol{B} = (\Phi(\boldsymbol{x}_t, \boldsymbol{x}_1), \Phi(\boldsymbol{x}_t, \boldsymbol{x}_1), \dots, \Phi(\boldsymbol{x}_t, \boldsymbol{x}_N)) \quad (2)$$

$$\boldsymbol{C}_Y(i,j) = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (3)$$

$$A = \Phi(\boldsymbol{x}_t, \boldsymbol{x}_t) \quad (4)$$

$$\Phi(\boldsymbol{x}_i, \boldsymbol{x}_j) = s \cdot \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\omega_d(\boldsymbol{x}_i(d) - \boldsymbol{x}_j(d))^2\right) + v \cdot \delta_{ij} \quad (5)$$

where D refers to the dimension of model input and δ_{ij} refers to the Kronecker delta. The hyperparameters could be denoted as $\boldsymbol{\theta} = (s, v, \omega_1, \dots, \omega_D)^T$. Then, identification of most probable hyperparameters $\boldsymbol{\theta}^*$ can be performed by maximizing the log-likelihood function for a GP model, as shown in the following equation:

$$\ln P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}^*) = -\frac{1}{2}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{Y}}^{-1}\boldsymbol{Y} - \frac{1}{2}\ln|\boldsymbol{C}_{\boldsymbol{Y}}| - \frac{N}{2}\ln 2\pi. \quad (6)$$

In wind power forecasting, the time-series model ignores the effect of exogenous inputs such as weather information and therefore could be considered to be time varying. It then becomes inappropriate to assume that all the historical data follow one joint Gaussian in GP. Moreover, the covariance matrix inversion in the GP implementation is computationally quite demanding, a variant of the standard GP is therefore required.

B. Persistence Model

Persistence model, also called the naïve predictor, is the most popular reference model, to compare the forecasting performance of any advanced models. It assumes that the generation at any time ahead $P(t+k)$ equals the production it has now $P(t)$. For short horizon prediction, this model works well due to the continuity of low-pressure atmosphere system. The model can be expressed as follows [2]:

$$\hat{P}(t+k) = P(t). \quad (7)$$

III. Temporally Local Gaussian Process

A. Model Description

A temporally local Gaussian process (TLGP) is developed employing only nearby local datasets and a "moving window" technique. TLGP differs from standard GP in two aspects. 1) TLGP defines a local window and uses only temporally local data inside the window to make predictions. 2) TLGP employs "moving window" technique to predict every training data, and further trains the model with the least-square technique. As the effective data window for forecasting each training data moves forward in time domain, the training data are estimated one by one and the sum of square error (SSE) could be obtained. Then, hyperparameters are optimized with least SSE techniques instead of maximum likelihood methods in GP. Therefore, TLGP is global in learning procedure while local in prediction process.

The local consecutive data in the effective window are denoted as $\boldsymbol{Y}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-M})^{\mathrm{T}}$ instead of $\boldsymbol{Y}$ suggesting that the effective window is adapting to time. Then, the standard GP is transformed into TLGP shown from (8) to (13)

$$\hat{y}_t = \boldsymbol{B}_t\boldsymbol{C}_t^{-1}\boldsymbol{Y}_t = \boldsymbol{B}_t\boldsymbol{C}_t^{-1}\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-M} \end{pmatrix} \quad (8)$$

$$\sigma^2 = A_t - \boldsymbol{B}_t{\boldsymbol{C}_t}^{-1}{\boldsymbol{B}_t}^T \quad (9)$$

$$\boldsymbol{B}_t = (\Phi(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}), \Phi(\boldsymbol{x}_t, \boldsymbol{x}_{t-2}), \dots, \Phi(\boldsymbol{x}_t, \boldsymbol{x}_{t-M})) \quad (10)$$

$$\boldsymbol{C}_t = \begin{bmatrix} \Phi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-1}), & \dots, & \Phi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-M}) \\ \vdots & \ddots & \vdots \\ \Phi(\boldsymbol{x}_{t-M}, \boldsymbol{x}_{t-1}), & \dots, & \Phi(\boldsymbol{x}_{t-M}, \boldsymbol{x}_{t-M}) \end{bmatrix} \quad (11)$$

$$A_t = \Phi(\boldsymbol{x}_t, \boldsymbol{x}_t) \quad (12)$$

$$\boldsymbol{x}_{t-i} = (y_{t-i-1}, y_{t-i-2}, \dots y_{t-i-L})^T. \quad (13)$$

In comparison with standard GP, both the cross-covariance vector $\boldsymbol{B}_t$ and self-covariance matrix $\boldsymbol{C}_t$ here are denoted with an extra t suggesting that for every new prediction point,

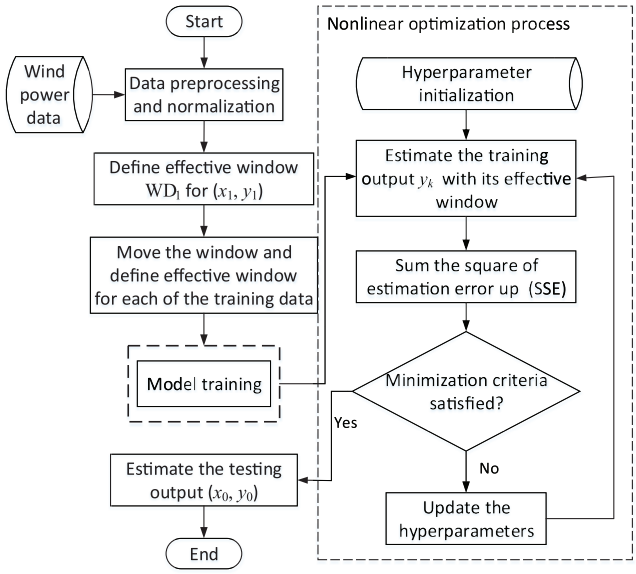

Fig. 1. Flowchart of TLGP and nonlinear optimization process.

the related effective data window is updated. After the SSE of trainin data has been obtained, the least-square error techniques such as genetic algorithm could be employed to get the optimal hyperparameters as shown in (14). Fig. 1 gives a flowchart for the proposed TLGP method embedded with the nonlinear optimization process. The optimization process automatically terminates if the maximum iteration number has been approached or the tolerance is below a preset threshold

$$\boldsymbol{\theta}^* = \arg * \min_{\boldsymbol{\theta}} \sum_{t=M+1}^{N} (\hat{y}_t - y_t)^2. \tag{14}$$

Overall, given a sequence of measurments $\{y_k\}_1^N$, the width of local window M, and the order of time lag $\boldsymbol{L}$, the TLGP can be summarized as follows to predict the output y_t at time t.

Step 1) Determine the useful data in the local widow of each training point y_k. Let $\{y_{kj}\}_{j=1}^M$ denote the local data in the kth window.

Step 2) Predict $\{y_k\}_1^N$ using its local data $\{y_{kj}\}_{j=1}^M$ with (8).

Step 3) Minimize the SSE in (14) and get the optimal hyperparameters $\boldsymbol{\theta}^*$.

Step 4) Determine the useful data in the local window of y_t, and forecast with the optimal hyperparameters using (8).

B. Computation Analysis

From the descriptionabove, the computational complexity of the TLGP and GP could be derived. The implementation of standard GP involves inversion of the global covariance matrix, which costs $O\left(N^3\right)$ flops in both the posterior distribution and the likelihood function. However, in the proposed method, the computational effort can be greatly reduced. In the cost function (14), the inversion of an M size matrix is carried out $N - M$ times, so the computation demand is $O\left((N-M) * M^3\right)$. Considering $(M \ll N)$, the computation is still greatly reduced. On the other hand, in the inference function (8) and the uncertainty estimation function (9), the computation cost become $O\left(M^3\right)$ which is obviously much smaller than that of the standard GP. The comparison is also given in Table I.

TABLE I
STANDARD GP AND TLGP COMPUTATIONAL COMPLEXITY

Computation	Learning	Inference	Uncertainty
Standard GP	$O(N^3)$	$O(N^3)$	$O(N^3)$
TLGP	$O(N * M^3)$	$O(M^3)$	$O(M^3)$

C. Iterative Multistep Forecasting

A multistep forecasting could generate multihorizon forecasting results for real-time balancing, control, and market planning. For iterative forecasting in a time-series system, after the new prediction is obtained, it is used to construct the new input and make the next step predictions [18]. k-step ahead forecasting of a discrete nonlinear system can be performed by repeating one-step ahead predictions. The state vector $\boldsymbol{x}_{t+k}$ corresponding to $\hat{y}_{t+k}$ is constructed by the previous estimates $(\hat{y}_{t+k-1}, \ldots, \hat{y}_{t+k-L})$, and the real measurement $(y_{t+k-L-1}, \ldots, y_{t-L})$. In this way, multistep forecasting scrolls forward.

The effectiveness of the proposed TLGP method is guranteed from three aspects.

1) For a sequence of N training samples, every sample is used in developing the model, and the sum of squared error of the N training samples is minimized to find the optimal hyperparameters in the model. Thus, the global property of the model training and the optimaility of the TLGP gurantee the accuracy of the obtained model, which is also verified in Section V.
2) For every new forecast, only its previous M local data $(M \ll N)$ that are highly correlated to the new point are employed. Such local mechanism in model prediction makes the model adaptive to the time-varying characteristic of the wind power system, which is further verified by the error distribution analysis in Section VI.
3) The proposed model provides a very efficient way for wind power forecasting in reducing the overall computation complexity, as being analyzed in the previous section.

IV. CASE STUDIES

In this section, the presented methods are applied to two wind farms labeled "A" and "B" located in USA and Ireland, respectively. Those wind farms differ from each other greatly on the installed capacity, one of which with approximately 300 MW and the other with 60 MW. Therefore, the effectiveness of the proposed analysis method on those examples could highlight the significance of the proposed method. For "A," data were sampled every second on August 2006 and then hourly average

value is calculated. Wind farm "B" is located in Galway. The Single wholesale Electricity Market (SEM) collected wind generation on October 2008 every 15 min and those data were averaged hourly for further application. The first two weeks' data were used for model training with TLGP to predict the output of the last two weeks. Hence, the last two weeks' measurements serve as testing data. Multihorizons forecasting from 1 to 12 h ahead were implemented with iterative multistep forecasting. As the choice of 1 to 12 h ahead forecasting is more beneficial for load dispatch and unit commitment, we do not consider the forecasting of shorter periods like 10 or 15 min. Generally speaking, persistence model is always more favorable for forecasting of very short periods than any other models due to its nature.

All these hourly averaged data were normalized over the individual full capacity, and then the squared exponential covariance function was employed with the mean function set as zero. Genetic algorithm was utilized to identify those hyperparameters in covariance function and the global optima could be approximately approached with multisimulations and proper settings of initial points. The selection of the parameter settings of TLGP, such as the number of observations in inputs (state vector $\boldsymbol{L}$) and the number of observations in effective window (window width M), was determined by trail-and-error experiments.

In order to evaluate the optimal settings, three performance metrics, namely normalized root-mean-square error (RMSE), normalized mean absolute error (MAE), and normalized mean absolute scaled error (MASE), were selected as shown in (15), (16), and (17), respectively. MAE is a linear score measuring the average absolute forecasting error, whereas RMSE is a quadratic score, better reflecting the error spikes. For MASE, the forecasting error is divided by the average change at two consecutive outputs; thus, it measures the error ratio over the average trend. For multistep forecasting, the trained model functions through all horizons, so the average normalized metrics with respect to the whole horizons were obtained and employed as the evaluating criteria. In addition to the multistep average errors, one-step forecasting errors were also calculated to reflect the effect of parameters settings

$$\text{RMSE} = \frac{1}{P_n}\sqrt{\sum_{i=1}^{N}\left(\frac{\hat{y}_i - y_i}{N}\right)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{P_n}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{N} \tag{16}$$

$$\text{MASE} = \frac{1}{P_n * N}\sum_{i=1}^{N}\left(\frac{|y_i - \hat{y}_i|}{\frac{1}{N-1}\sum_{i=2}^{N}|y_i - y_{i-1}|}\right) \tag{17}$$

where P_n is the nominal capacity of the particular wind farm.

A. Site "A" in USA

For wind farm "A," 36 simulations have been carried out for one single wind farm with $\boldsymbol{L} \in [1, 6]$ and $M \in [2, 7]$. Here, M starts from 2 to guarantee more information are employed than that of the persistence model.

For each time lag $\boldsymbol{L}$ in $[1, 6]$, the corresponding optimal widow size M was selected with respect to the multistep average error. The result is shown in Fig. 2(a). As both MAE and MASE are linear scores of absolute forecasting error, they follow the same trend. Thus, only the RMSE and MAE results are shown. It can be seen that the optimal solutions for M are around 3 or 4 h for any $\boldsymbol{L}$ and the multistep average RMSE and MAE reach the minimum value at the same point. It shows that RMSE and MAE have similar trend no matter for multistep or one step. However, for one fixed metric such as RMSE, the one-step performance shows different trend with that of multistep average. In (a), the minimum multistep error lies in (2, 3).

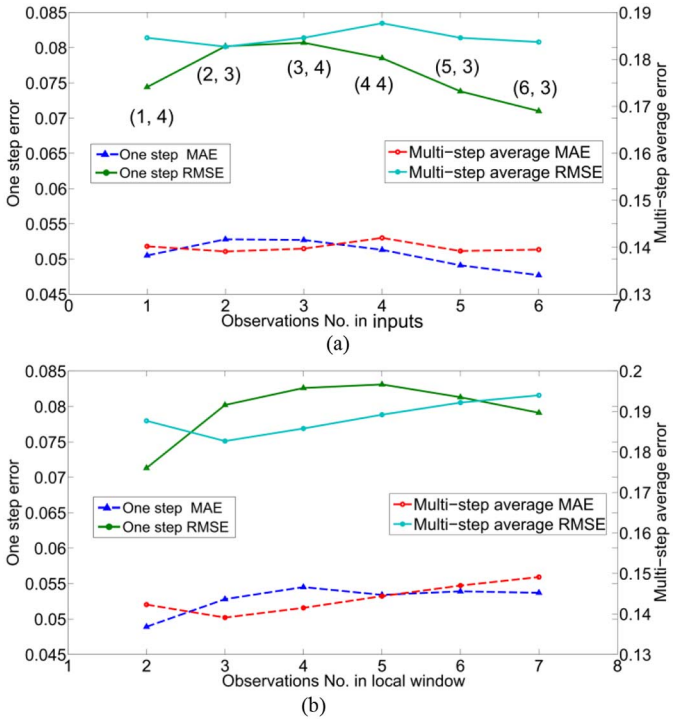

Fig. 2. Dependence of forecasting errors on time lag and local window size. (a) Order of time lag. (b) Width of local window.

Fig. 2(b) shows the forecasting errors with fixed $\boldsymbol{L} = 2$. Still, RMSE and MAE show similar trend, and the first-step forecasting and multistep average optimizes at different settings. The best one-step forecasting occurs at $M = 2$, while optimal multistep average optimal lies in $M = 3$.

Fig. 3 shows the wind generation measurements and forecasting at optimal one-step forecasting (3, 2). It can be seen that the measured data show high variability and ramping events occur at some point which increase the forecasting difficulty. Although the MAE is only 5% of the installed capacity P_n, the maximum absolute error could reach 40%. Such moments happens at ramping events which are quite common in wind power forecasting.

B. Site "B" in Ireland

Similarly, the trial-and-error method has been applied to wind farm "B" to select the optimal model parameters. Seventy-two groups of parameters are set up in simulation and the forecasting metrics are obtained. The first four rows in Table II compared the results of some settings to select the optimal models with respect to one-step error, and the last four rows are

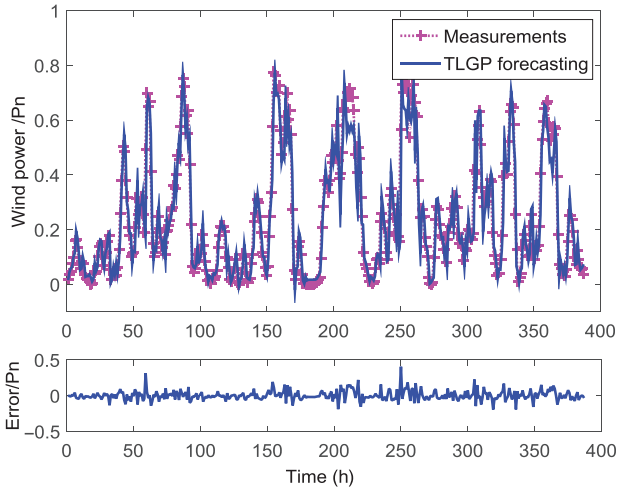

Fig. 3. One-hour forecasting result of wind farm "A" in USA.

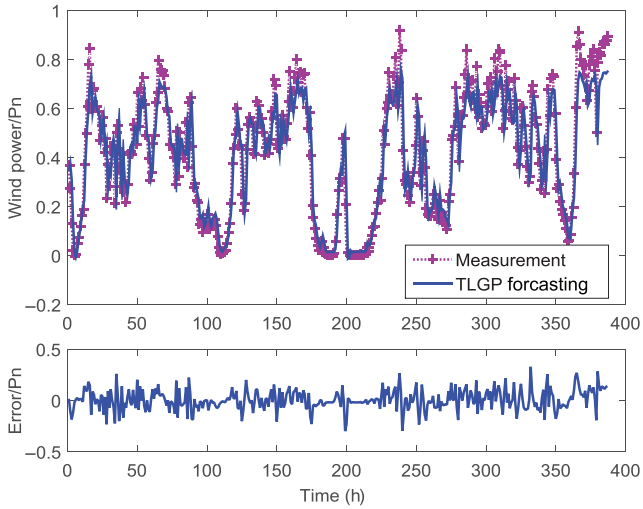

Fig. 4. One-hour forecasting result of wind farm "B" in Ireland.

TABLE II
OPTIMAL MODEL PARAMETERS FOR TLGP IN WIND FARM "B"

(L, M)	(3, 2)	(5, 2)	**(7, 2)**	(9, 2)	(11, 2)	(13, 2)
One step MAE	0.0644	0. 2360	**0.0616**	0.0628	0.0640	0.0660
One step RMSE	0.0868	0.0860	**0.0852**	0.0860	0.0868	0.0888
One step MASE	0.8663	0.8674	**0.8371**	0.8454	0.8477	0.8495
(L,M)	(2, 3)	(3, 2)	(4, 3)	(5, 4)	**(6, 4)**	(7, 3)
Multistep average MAE	0.1476	0.1488	0.1488	0.1444	**0.1440**	0.1472
Multistep average RMSE	0.1936	0.1972	0.1968	0.1908	**0.1896**	0.1936
Multistep average MASE	1.9798	2.1015	2.0296	1.9492	**1.9391**	1.9884

about multistep average errors. As is shown, the optimal forecasting with minimum one-step error including RMSE, MAE, and MASE occurs at $\boldsymbol{L} = 7$ and $M = 2$, and the optimal multistep forecasting with respect to the average of those metrics lies under (6, 4). The one-step forecasting tends to employ fewer observations in local window while require more elements in inputs in comparison with the multistep one. Multistep error considers the performance over the whole horizon, which potentially utilizes longer local window width to meet the middle horizon requirements. Moreover, the three metrics tend to show similar trend.

Under the optimal one-step settings, the forecasting results in wind farm "B" are shown in Fig. 4. The mean wind generation is around 40% of the installed capacity. Although the maximum forecasting error could reach 32% of the capacity, the mean abosolute forecasting error is only about 6.5% of the capacity. It can be seen that the most obvious amplitute error of forecasting happens at the spike points of wind generation. Besides amplitute error, phase error makes the other important component of the main error forms, which describes the time lag between forecasting and real generation especially at the ramping events. Such kind of error makes it impossible to compare forecasting performance quantitively between different wind farms, but in one particular wind farm, the comparison between different methods could be carried out, which is illustrated in the next section.

V. RESULTS ANALYSIS AND DISCUSSION

With the forecasting results from the previous session, the metrics of TLGP in wind farm "A" in comparison with those of standard GP, ARMA, RBF, and persistence model are analyzed in Fig. 5. With the parameters set as (3, 4), TLGP outperforms other benchmark models in most of the horizons in terms of MAE, RMSE, and MASE. As the multistep forecasting parameters are selected according to the minimum average error, there are no guaranteed improvements over all horizons. Consequently, it is not surprising that the first step prediction shows minor advantage over other model under such settings, and the maximum improvement over GP occurs at the middle horizon. Similarly, for the longer horizon forecasting, the number of observations in local window needs to be larger to get better forecasting results and improvements.

Table III shows the forecasting improvements of TLGP over other benchmark models with respect to different metrics. Those results look quite optimistic with average improvement over persistence more than 18% and over GP more than 5%. The average improvement over GP with respect to MAE and MASE is almost twice that of RMSE, and from Fig. 5, after 12 h, the improvement in terms of RMSE will disappear. Such phenomenon suggests that TLGP is quite capable of capturing the relatively stable local generation window but shows less advantage in minimizing the ramping event forecasting error in comparison with GP. However, the result is still acceptable with improved forecasting results. More importantly, the forecasting with GP takes 120 s, while TLGP just costs 40 s, which is a proof of saving computation resources. Furthermore, TLGP outperforms RBF and ARMA showing its advantages even outside the GP family.

In wind farm "B," similar results show up in Fig. 6. TLGP outperforms GP and persistence in most of the horizons under the parameters settings of (6, 4). The maximum improvement occurs in the middle horizon, and TLGP will be less satisfying after 12 h in term of RMSE. Table IV shows the quantitative improvement of TLGP over the benchmarks. The improvements over GP, ARMA, and RBF in "B" are greater than that

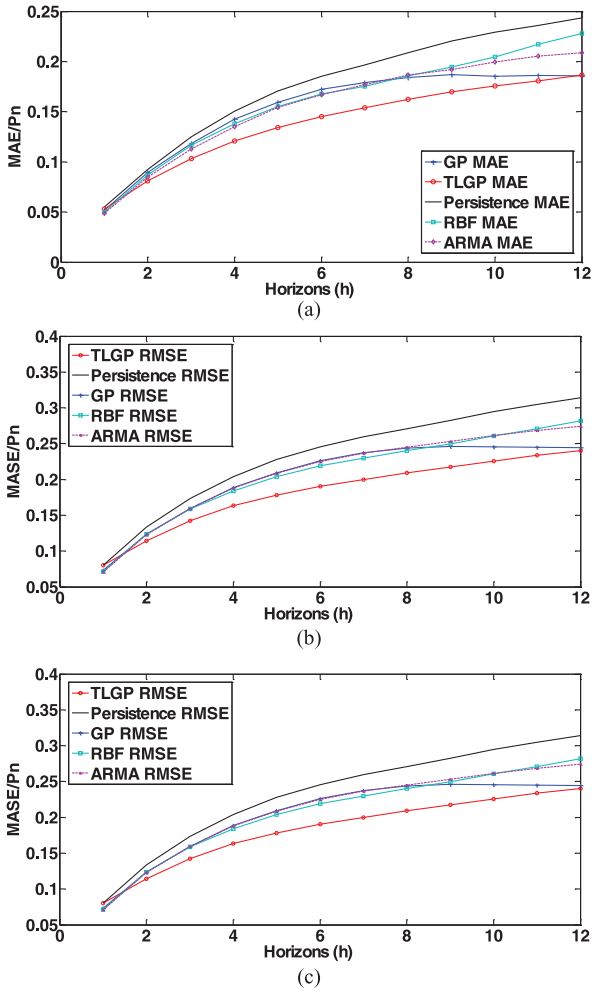

Fig. 5. Forecasting error comparison of wind farm "A" in USA. (a) MAE. (b) RMSE. (c) MASE.

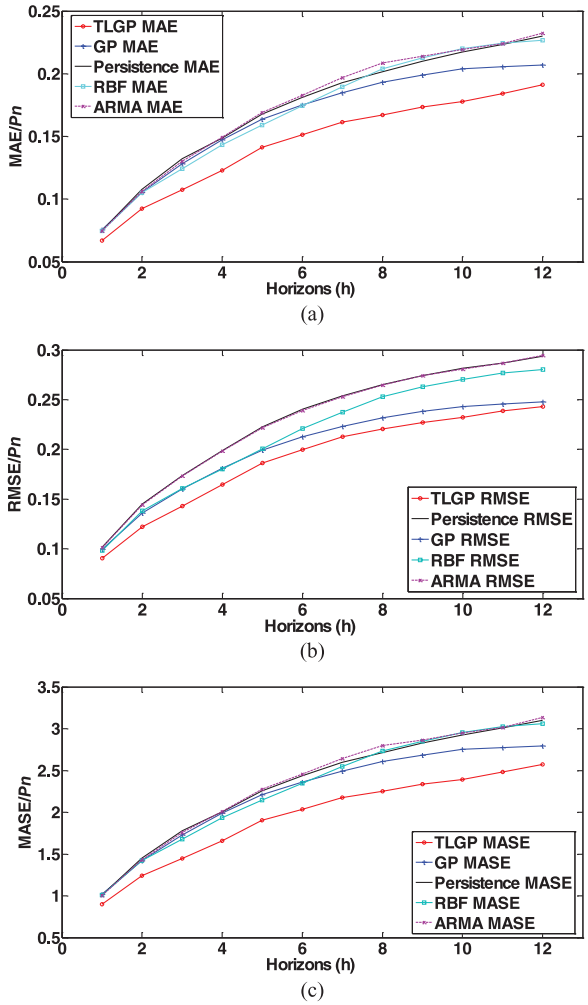

Fig. 6. Forecasting error comparison of wind farm "B." (a) MAE. (b) RMSE. (c) MASE.

TABLE III
IMPROVEMENT OF TLGP OVER THE BENCHMARK MODELS IN WIND FARM "A"

Metrics improvement	Max RMSE (%)	Mean RMSE (%)	Max MAE (%)	Mean MAE (%)	Max MASE (%)	Mean MASE (%)
Persistence	23.7	18.5	23.9	18.9	23.4	19.3
GP	9.97	5.49	14.9	10	16.2	10.9
RBF	14.7	10.2	17.9	11.3	16.6	10.5
ARMA	15.4	10.9	12.8	9.2	12.1	8.4

TABLE IV
IMPROVEMENT OF TLGP OVER THE BENCHMARK MODELS IN WIND FARM "B"

Metrics improvement	Max RMSE (%)	Mean RMSE (%)	Max MAE (%)	Mean MAE (%)	Max MASE (%)	Mean MASE (%)
Persistence	18.39	16.49	19.1	16.6	19.1	16.9
GP	11.59	6.5	17.37	13.21	18	13.6
RBF	14.0	11.2	19.2	14.9	19.0	14.9
ARMA	17.6	16.2	19.8	16.9	19.5	16.9

in "A," while the skill score over persistence is less than that in "A." This suggests that for such wind farm with smaller capacity, the performance drops of other advanced models are more obvious than that of TLGP. In other words, TLGP shows more advantage over those advanced benchmarks for power forecasting of smaller wind farms.

VI. PROBABILISTIC ANALYSIS OF ERROR DISTRIBUTION

A. Statistical Analysis of Error Distribution

Although, in those examples, the proposed TLGP has shown its advantages, those results are not enough to guarantee that TLGP will outperform other models regardless of different wind farms at various time of the year. For farms with more condensed severe ramping events in a particular time of the year, the result would be less satisfactory. For example, taking one wind farm in Ireland, the hourly wind generation varies fast as shown in Fig. 7. In such case, the TLGP would not beat other models over all horizons. In order to examine the suitability of the proposed method, the distribution probability of forecasting residuals is analyzed.

The ideal model for forecasting a time-series wind power output is expected to generate regular residuals that follow Gaussian distribution, while any nonlinearity or non-Gaussian distribution in the process is either captured by the model or eliminated through modeling. Further, the smaller the variance, the better quality the forecasting results. In this part, 1500 outputs from Fig. 7 are predicted with five methods,

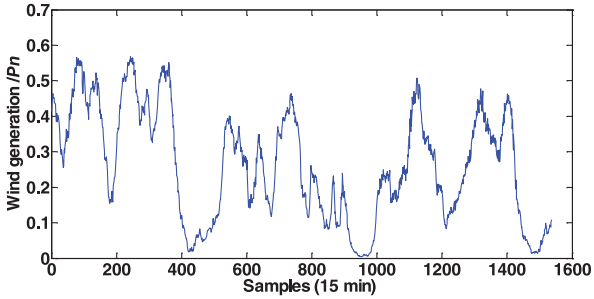

Fig. 7. Wind generation in a wind farm of Ireland.

respectively, and the generated residuals for each horizon forecasting are divided into 30 even intervals according to their range. The number of residuals in each interval could be statistically counted and a bar chart could be sketched. In this way, the probability of the residuals falling on each interval could be calculated and the profile of the overall chart will reflect the distribution of forecasting errors. Fig. 8 shows the analysis result of 3-h ahead forecasting. The forecasting with persistence and ARMA model in (a) and (d) are most unlike Gaussian distribution. The error stay in the range of $(-0.2,\ 0.2)$, but the distribution is severely asymmetric and the profiles do not follow the shape of Gaussian distribution. Such shortages are improved in the forecasting by GP and RBF resorting to the more symmetric bars in (b) and (c). However, at some parts of the profiles, the probability changes linearly rather than exponentially, which shows the flaws of such forecasting. In (e), TLGP generates some large errors in forecasting some ramping events. Although those errors approach -0.5 in the horizontal axis, TLGP makes the most asymmetric forecasting in the most probable confidence region from $(-0.2,\ 0.2)$. Moreover, the peak probability of (e) is approaching 15%, which is an exponential rise from the intervals beside it.

Those comparisons and the analysis show that the profile of TLGP forecasting residuals follows Gaussian distribution better. Such result proves the effectiveness of the proposed concept of "moving window": the new output will follow one joint Gaussian distribution with the local window rather than the total available generation data. In the next section, such statement will be verified quantitatively.

B. Quantitative Verification of Error Distribution

Skewness and kurtosis are two main metrics to evaluate the shape of variable distribution. While skewness measures whether the distribution is symmetric, kurtosis evaluates how tall and sharp the central peak is. For the distributions with zero mean, the metrics can be expressed as follows:

$$s = E\left(\varepsilon^3\right)/\sigma^3 \tag{18}$$

$$k = E\left(\varepsilon^4\right)/\sigma^4 \tag{19}$$

where σ refers to the standard deviation and ε denotes forecasting error.

For the errors at different forecasting horizons, both metrics are calculated and the results are shown in Table V. If skewness is positive, the data are positively skewed meaning that the

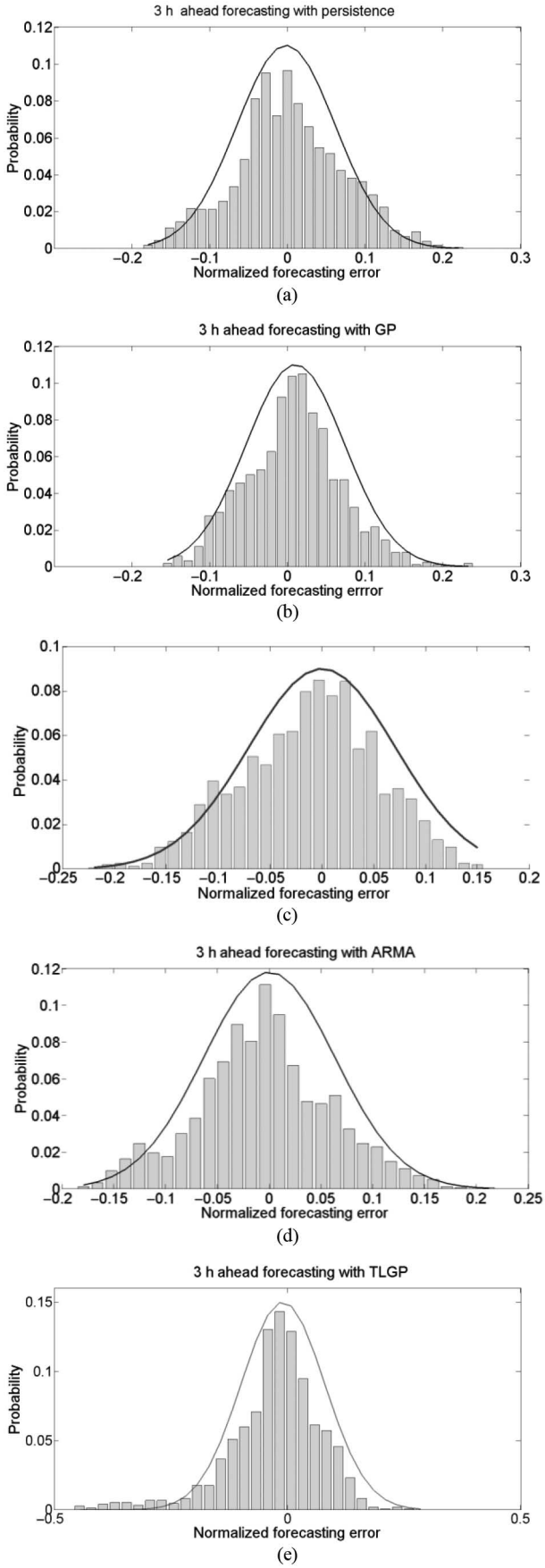

Fig. 8. Forecasting error distribution with five separate methods. (a) Persistence. (b) GP. (c) RBF. (d) ARMA. (e) TLGP.

TABLE V
SKEWNESS OF ERROR DISTRIBUTION PREDICTED WITH DIFFERENT METHODS

Horizons (h)	2	4	6	8	10
Persistence	0.128	0.151	0.176	0.184	0.208
GP	−0.029	−0.028	−0.049	−0.078	−0.095
TLGP	−0.167	−0.025	0.011	−0.019	0.008
RBF	−0.0806	−0.0675	−0.0504	−0.0170	0.0119
ARMA	0.0994	0.1187	0.1355	0.1386	0.1609

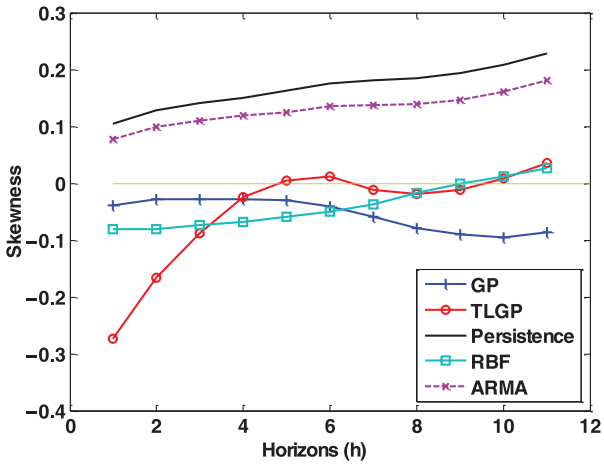

Fig. 9. Skewness of residuals at varied horizons predicted with five methods.

TABLE VI
KURTOSIS OF ERROR DISTRIBUTION PREDICTED WITH DIFFERENT METHODS

Methods	2	4	6	8	10
Persistence	2.74	2.74	2.76	2.74	2.64
GP	2.54	2.41	2.28	2.20	2.12
TLGP	3.05	2.92	2.86	2.88	2.90
RBF	2.4275	2.3540	2.2993	2.2756	2.2940
ARMA	2.8006	2.7900	2.7738	2.7494	2.6587

right tail of the distribution is longer than the left. Otherwise, it is called negatively skewed. If skewness is between -0.5 and $+0.5$, the distribution is approximately symmetric. It can be found that for those three methods, the error distributions are about symmetric at all horizons. In Fig. 9, the skewness trends of five methods with respect to horizons are sketched. In the first three steps, TLGP loses to the other four due to bigger skewness (reflected by long left tail in Fig. 8), but it returns and stabilizes around the desired zero skewness in the following horizons showing its better ability in producing symmetric errors.

The kurtosis of the obtained residual distribution is shown in Table VI. A normal distribution has a kurtosis of exact 3. So, the errors at different horizons all follow a near-Gaussian distribution. In Fig. 10, it shows the comparison of kurtosis produced by five different methods at different horizons. The finding is that kurtosis of TLGP residuals is the nearest to that of standard Gaussian distribution denoted with dashed line.

Therefore, the probabilistic analysis of forecasting error is obtained. Such results can be applied further to the energy storage system design and economic marketing. The skewness and kurtosis comparison shows that the error produced by TLGP is most likely Gaussian. Such analysis reflects that the proposed TLGP fits best to the wind power modeling.

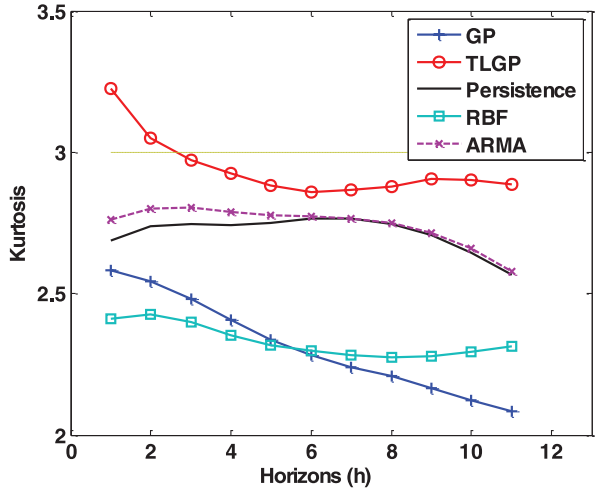

Fig. 10. Kurtosis of residuals at varied horizons predicted with five methods.

VII. CONCLUSION

In this paper, a hybrid deterministic–probabilistic wind power forecasting method TLGP has been developed to adapt to the time-varying characteristic of wind power, to overcome the computation difficulty of the standard GP, and to generate statistic error distribution results. The effectiveness of the proposed method has been verified by its application to two wind farms in USA and Ireland, respectively. The multistep forecasting results from 1 to 12 h have been analyzed and compared with four benchmark models in terms of three metrics (RMSE, MAE, and MASE). Moreover, the forecasting residuals with these methods are analyzed to check the deviation from standard Gaussian distribution.

It has been demonstrated that the proposed hybrid deterministic–probabilistic method reduces the computational complexity during the learning and inference process compared to the standard GP. Moreover, the proposed approach shows great effectiveness at overcoming time-varying characteristic at the two wind farms by producing a smaller forecasting error than those of other models. Although it displays less advantage in forecasting ramp events, the statistical error analysis shows that TLGP produces residuals that are most likely Gaussian. This finding reveals an important advantage of TLGP from a probabilistic standpoint.

ACKNOWLEDGMENT

J. Yan would like to thank the Chinese Scholarship Council (CSC) and U.K.–China Science Bridge Project for their sponsorship during her Ph.D. study at Queen's University Belfast and A. M. Foley also thanks EirGrid for use of datasets from the SEM.

REFERENCES

[1] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012.

[2] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind power: A literature overview," *ANEMOS.plus*, Risø DTU, Wind Energy Division, Tech. Rep. Roskilde, Denmark, 2011 [Online]. Available: http://orbit.dtu.dk/fedora/objects/orbit:83397/datastreams/file_5277161/content.

[3] R. J. Bessa *et al.*, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: A case study," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 827–836, Oct. 2012.

[4] A. Botterud *et al.*, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 250–261, Jan. 2013.

[5] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[6] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.

[7] N. Amjady, F. Keynia, and H. Zareipour, "Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization," *IEEE Trans. Sustain. Energy*, vol. 2, no. 3, pp. 265–276, Jul. 2011.

[8] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov. 2010.

[9] W. Can, X. Zhao, P. Pinson, Y. Zhao, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[10] J. Kiviluoma *et al.*, "Short-term energy balancing with increasing levels of wind energy," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 769–776, Oct. 2012.

[11] C. Lowery and M. O'Malley, "Impact of wind forecast error statistics upon unit commitment," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 760–768, Oct. 2012.

[12] A. Shortt, J. Kiviluoma, and M. O'Malley, "Accommodating variability in generation planning," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 158–169, Feb. 2013.

[13] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 983–991, Aug. 2008.

[14] S. Tewari, C. J. Geyer, and N. Mohan, "A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2031–2039, Nov. 2011.

[15] M. Lange, "On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors," *J. Sol. Energy Eng.*, vol. 127, no. 2, pp. 177–184, 2005.

[16] P. Kou, F. Gao, and X. Guan, "Sparse online warped Gaussian process for wind power probabilistic forecasting," *Appl. Energy*, vol. 108, pp. 410–428, 2013.

[17] J. Yan, Z. Yang, K. Li, and Y. Xue, "A variant Gaussian process for short-term wind power forecasting based on TLBO," *Communications in Computer and Information Science*, vol. 463, pp. 165–174, 2014.

[18] A. Girard, C. E. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith, "Multiple-step ahead prediction for non-linear dynamic systems—A Gaussian process treatment with propagation of the uncertainty," *Adv. Neural Inf. Process. Syst.*, vol. 15, pp. 529–536, 2003.

Juan Yan (S'13–M'15) received the B.Eng. and M.Eng. degrees in electrical engineering from Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively. She is currently pursuing the Ph.D. degree in electrical engineering at the Group of Energy, Power, and Intelligent Control, Queen's University of Belfast, Belfast, U.K.

Her research interests include machine learning techniques such as Gaussian process and neural networks, uncertainty analysis, nonlinear optimization algorithms, computational complexity reduction techniques, and their applications in wind power forecasting.

Kang Li (M'05–SM'11) received the B.Sc. degree in industrial automation from Xiangtan University, Xiangtan, China, in 1989, the M.Sc. degree in control theory and applications from Harbin Institute of Technology, Harbin, China, in 1992, the Ph.D. degree in control theory and applications from Shanghai Jiaotong University, Shanghai, China, in 1995, and the D.Sc. degree in science from Queen's University Belfast, Belfast, U.K., in 2015.

He is currently a Professor in Intelligent Systems and Control with the School of Electronics, Electrical Engineering, and Computer Science, Queen's University Belfast. His research interests include nonlinear system modeling, identification and control, and bioinspired computational intelligence, with applications to energy and power systems, renewable energies, smart grid, electric vehicles, polymer processing, the development of advanced control technologies for decarbonizing the entire energy systems from top to tail, and the development of a new generation of low-cost minimally invasive monitoring system and intelligent control platforms for energy-intensive industries.

Er-Wei Bai (M'90–SM'00–F'04) received the B.S. degree in physics from Fudan University, Shanghai, China, in 1977, the M.S. degree in electrical engineering and computer sciences from Shanghai Jiaotong University, Shanghai, China, in 1982 and the Ph.D. degree in electrical engineering and computer sciences from the University of California at Berkeley, Berkeley, USA, in 1987.

He is currently a Professor of Electrical Engineering with the University of Iowa, Iowa City, IA, USA, where he teaches and conducts research in the areas of identification, control, signal processing, and their applications in engineering and medicine.

Dr. Bai was the recipient of the Regents Award for Faculty Excellence as well as the President's Award for Teaching Excellence.

Jing Deng received the M.Sc. degree in control theory and control engineering from Shanghai University, Shanghai, China, in 2008, and the Ph.D. degree in electrical engineering from Queen's University Belfast, Belfast, U.K., in 2011.

He then started to work as a Postdoctoral Research Fellow with Queen's University Belfast for a few projects, including process monitoring and intelligent control for polymer processing and smart charging for electric vehicles. His research interests include nonlinear system modeling, optimization, and fuzzy control with their applications to smart grid, electric vehicles, and renewable energies.

Aoife M. Foley (M'09) received the B.E. (Hons.) degree in civil and environmental engineering from the University College Cork, Cork, Ireland, in 1996, the M.Sc. (Hons.) degree in transportation from Trinity College Dublin, Dublin, Ireland, in 1999, and the Ph.D. degree in energy engineering from the University College Cork, Cork, Ireland, in 2011.

She is a Lecturer with the School of Mechanical and Aerospace Engineering, Queen's University Belfast, Belfast, U.K. Her research interests include wind power forecasting, wind power integration using storage and demand response, smart grid technology, and electricity markets.

Dr. Foley is a Fellow of Engineers Ireland (2012).